CLAIMS

1. A method for related term suggestion, the method comprising:

generating term clusters as a function of calculated similarity of term vectors, each term vector being generated from search results associated with a set of high frequency of occurrence (FOO) historical queries previously submitted to a search engine; and

responsive to receiving a term/phrase from an entity, evaluating the term/phrase in view of terms/phrases in the term clusters to identify one or more related term suggestions.

2. A method as recited in claim 1, and wherein a multi-sense query comprises the term/phrase.

3. A method as recited in claim 1, and wherein the entity is a computer-program application and/or an end-user.

4. A method as recited in claim 1, further comprising determining the calculated similarity as follows:

$$sim(q_j, q_k) = \sum_{i=1}^{d} w_{ij} \cdot w_{ik} \; ;$$

wherein weight w for the ith vector's j^{th} term is calculated as follows:

$$w_{ij} = TF_{ij} x \log(N / DF_j) \; ; \text{ and}$$

wherein TF_{ij} represents term frequency, N is a total number of query terms, and DF_j is a number of extracted feature records that contain term j.

5. A method as recited in claim 1, further comprising:

collecting historic query terms from a query log; and

determining ones of the historic query terms with a high FOO.

6. A method as recited in claim 1, further comprising before creating the term clusters:

reducing dimensionality of the term vectors; and

normalizing the term vectors.

7. A method as recited in claim 1, wherein evaluating further comprises:

identifying a match between the term/phrase and term(s)/phrase(s) from one or more term clusters; and

responsive to identifying, generating related term suggestion(s) comprising the term(s)/phrase(s).

8. A method as recited in claim 7, wherein the related term suggestion(s) further comprise for each term/phrase of the term(s)/phrase(s), a frequency of occurrence value indicating a number of times the term/phrase occurs in a set of mined historical queries.

9. A method as recited in claim 1, wherein generating the term clusters further comprises:

sending respective ones of the high FOO historical queries to the search engine to obtain the search results;

extracting features from at least a subset of search results corresponding to the respective ones; and

producing term vectors from the features as a function of term and inverted term frequencies.

10. A method as recited in claim 9, and wherein the features comprise a title, description, and/or context for the respective ones of the high FOO historical query terms.

11. A method as recited in claim 9, and wherein the respective ones comprise top ranked ones of the search results.

12.   A method as recited in claim 1, wherein the term clusters are a first set of term clusters, and wherein the method further comprises:

determining that there is no match between the term/phrase and the terms/phrases; and

responsive to the determining:

making a second set of term clusters from calculated similarity of term vectors, each term vector being generated from search results associated with a set of low FOO historical queries previously submitted to the search engine; and

evaluating the term/phrase in view of terms/phrases of the second set of term clusters to identify one or more related term suggestions.

13.   A method as recited in claim 12, wherein making further comprises:

identifying the low FOO historical queries from historical queries mined from a query log;

sending respective ones of at least a subset of the low FOO historical queries to the search engine to obtain search results;

extracting features from at least a subset of search results; and

producing the term vectors from the features as a function of term and inverted term frequencies

14. A method as recited in claim 13, and further comprising after clustering:

determining that there is no match between the term/phrase and term(s)/phrase(s) from the first set of term clusters, the first set being based on high FOO historical queries; and

responsive to the determining, identifying a match between the term/phrase and term(s)/phrase(s) from one or more of the second set of term clusters, the second set being based on low FOO historical queries; and

responsive to identifying, generating related term suggestion(s) comprising the term(s)/phrase(s).

15. A computer-readable medium comprising computer-executable instructions for:

generating term clusters as a function of calculated similarity of term vectors, each term vector being generated from search results associated with a set of high frequency of occurrence (FOO) historical queries previously submitted to a search engine; and

responsive to receiving a term/phrase from an entity, evaluating the term/phrase in view of terms/phrases in the term clusters to identify one or more related term suggestions.

16. A computer-readable medium as recited in claim 15, and wherein a multi-sense query comprises the term/phrase.

17.　　A computer-readable medium as recited in claim 15, and wherein the entity is a computer-program application and/or an end-user.

18.　　A computer-readable medium as recited in claim 15, further comprising computer-executable instructions for determining the calculated similarity as follows:

$$sim(q_j, q_k) = \sum_{i=1}^{d} w_{ij} \cdot w_{ik} \; ;$$

wherein weight w for the ith vector's j^{th} term is calculated as follows:

$$w_{ij} = TF_{ij} x \log(N / DF_j) \; ; \text{ and}$$

wherein TF_{ij} represents term frequency, N is a total number of query terms, and DF_j is a number of extracted feature records that contain term j.

19.　　A computer-readable medium as recited in claim 15, further comprising computer-executable instructions for:

　　　　collecting historic query terms from a query log; and

　　　　determining ones of the historic query terms with a high FOO.

20.　　A computer-readable medium as recited in claim 15, before creating the term clusters, further comprising computer-executable instructions for:

　　　　reducing dimensionality of the term vectors; and

　　　　normalizing the term vectors.

21. A computer-readable medium as recited in claim 15, wherein evaluating further comprises computer-executable instructions for:

identifying a match between the term/phrase and term(s)/phrase(s) from one or more term clusters; and

responsive to identifying, generating related term suggestion(s) comprising the term(s)/phrase(s).

22. A computer-readable medium as recited in claim 21, wherein the related term suggestion(s) further comprise for each term/phrase of the term(s)/phrase(s), a frequency of occurrence value indicating a number of times the term/phrase occurs in a set of mined historical queries.

23. A computer-readable medium as recited in claim 15, wherein generating the term clusters further comprises computer-executable instructions for:

sending respective ones of the high FOO historical queries to the search engine to obtain the search results;

extracting features from at least a subset of search results corresponding to the respective ones; and

producing term vectors from the features as a function of term and inverted term frequencies.

24. A computer-readable medium as recited in claim 23, and wherein the features comprise a title, description, and/or context for the respective ones of the high FOO historical query terms.

25. A computer-readable medium as recited in claim 23, and wherein the respective ones comprise top ranked ones of the search results.

26. A computer-readable medium as recited in claim 15, wherein the term clusters are a first set of term clusters, and wherein the computer-executable instructions further comprise instructions for:

determining that there is no match between the term/phrase and the terms/phrases; and

responsive to the determining:

making a second set of term clusters from calculated similarity of term vectors, each term vector being generated from search results associated with a set of low FOO historical queries previously submitted to the search engine; and

evaluating the term/phrase in view of terms/phrases of the second set of term clusters to identify one or more related term suggestions.

27.   A computer-readable medium as recited in claim 26, wherein making further comprises computer-executable instructions for:

identifying the low FOO historical queries from historical queries mined from a query log;

sending respective ones of at least a subset of the low FOO historical queries to the search engine to obtain search results;

extracting features from at least a subset of search results; and

producing the term vectors from the features as a function of term and inverted term frequencies

28.   A computer-readable medium as recited in claim 27, and further comprising computer-executable instructions, after clustering, for:

determining that there is no match between the term/phrase and term(s)/phrase(s) from the first set of term clusters, the first set being based on high FOO historical queries; and

responsive to the determining, identifying a match between the term/phrase and term(s)/phrase(s) from one or more of the second set of term clusters, the second set being based on low FOO historical queries; and

responsive to identifying, generating related term suggestion(s) comprising the term(s)/phrase(s).

29.   A computing device comprising:

a processor; and

a memory coupled to the processor, the memory comprising computer-program instructions executable by the processor for:

generating term clusters as a function of calculated similarity of term vectors, each term vector being generated from search results associated with a set of high frequency of occurrence (FOO) historical queries previously submitted to a search engine; and

responsive to receiving a term/phrase from an entity, evaluating the term/phrase in view of terms/phrases in the term clusters to identify one or more related term suggestions.

30.   A computing device as recited in claim 29, and wherein a multi-sense query comprises the term/phrase.

31.   A computing device as recited in claim 29, and wherein the entity is a computer-program application and/or an end-user.

32.   A computing device as recited in claim 29, further comprising computer-executable instructions for determining the calculated similarity as follows:

$$sim(q_j, q_k) = \sum_{i=1}^{d} w_{ij} \cdot w_{ik} \; ;$$

wherein weight w for the ith vector's j^{th} term is calculated as follows:

$$w_{ij} = TF_{ij} x \log(N / DF_j) \; ; \text{ and}$$

wherein TF_{ij} represents term frequency, N is a total number of query terms, and DF_j is a number of extracted feature records that contain term j.

33.   A computing device as recited in claim 29, further comprising computer-executable instructions for:

collecting historic query terms from a query log; and

determining ones of the historic query terms with a high FOO.

34.   A computing device as recited in claim 29, before creating the term clusters, further comprising computer-executable instructions for:

reducing dimensionality of the term vectors; and

normalizing the term vectors.

35. A computing device as recited in claim 29, wherein evaluating further comprises computer-executable instructions for:

identifying a match between the term/phrase and term(s)/phrase(s) from one or more term clusters; and

responsive to identifying, generating related term suggestion(s) comprising the term(s)/phrase(s).

36. A computing device as recited in claim 35, wherein the related term suggestion(s) further comprise for each term/phrase of the term(s)/phrase(s), a frequency of occurrence value indicating a number of times the term/phrase occurs in a set of mined historical queries.

37. A computing device as recited in claim 29, wherein generating the term clusters further comprises computer-executable instructions for:

sending respective ones of the high FOO historical queries to the search engine to obtain the search results;

extracting features from at least a subset of search results corresponding to the respective ones; and

producing term vectors from the features as a function of term and inverted term frequencies.

38. A computing device as recited in claim 37, and wherein the features comprise a title, description, and/or context for the respective ones of the high FOO historical query terms.

39. A computing device as recited in claim 37, and wherein the respective ones comprise top ranked ones of the search results.

40. A computing device as recited in claim 29, wherein the term clusters are a first set of term clusters, and wherein the computer-executable instructions further comprise instructions for:

determining that there is no match between the term/phrase and the terms/phrases; and

responsive to the determining:

making a second set of term clusters from calculated similarity of term vectors, each term vector being generated from search results associated with a set of low FOO historical queries previously submitted to the search engine; and

evaluating the term/phrase in view of terms/phrases of the second set of term clusters to identify one or more related term suggestions.

41.   A computing device as recited in claim 40, wherein making further comprises computer-executable instructions for:

identifying the low FOO historical queries from historical queries mined from a query log;

sending respective ones of at least a subset of the low FOO historical queries to the search engine to obtain search results;

extracting features from at least a subset of search results; and

producing the term vectors from the features as a function of term and inverted term frequencies

42.   A computing device as recited in claim 41, and further comprising computer-executable instructions, after clustering, for:

determining that there is no match between the term/phrase and term(s)/phrase(s) from the first set of term clusters, the first set being based on high FOO historical queries; and

responsive to the determining, identifying a match between the term/phrase and term(s)/phrase(s) from one or more of the second set of term clusters, the second set being based on low FOO historical queries; and

responsive to identifying, generating related term suggestion(s) comprising the term(s)/phrase(s).

43.   A computing device comprising:

generating means to generate term clusters as a function of calculated similarity of term vectors, each term vector being generated from search results associated with a set of high frequency of occurrence (FOO) historical queries previously submitted to a search engine; and

responsive to receiving a term/phrase from an entity, evaluating means to evaluate the term/phrase in view of terms/phrases in the term clusters to identify one or more related term suggestions.

44.   A computing device as recited in claim 43, and wherein a multi-sense query comprises the term/phrase.

45.   A computing device as recited in claim 43, and wherein the entity is a computer-program application and/or an end-user.

46.   A computing device as recited in claim 43, and further comprising:

collecting means to collect historic query terms from a query log; and

determining means to determine ones of the historic query terms with a high FOO.

47.   A computing device as recited in claim 43, wherein the evaluating means further comprise:

identifying means to identify a match between the term/phrase and term(s)/phrase(s) from one or more term clusters; and

responsive to identifying, generating means to generate related term suggestion(s) comprising the term(s)/phrase(s).

48.   A computing device as recited in claim 43, wherein the generating means ti generate the term clusters further comprise:

sending means to send respective ones of the high FOO historical queries to the search engine to obtain the search results;

extracting means to extract features from at least a subset of search results corresponding to the respective ones; and

producing means to produce term vectors from the features.

49. A computing device as recited in claim 43, wherein the term clusters are a first set of term clusters, and wherein the computing device further comprises:

determining means to determine that there is no match between the term/phrase and the terms/phrases; and

responsive to the determining:

making means to make a second set of term clusters from calculated similarity of term vectors, each term vector being generated from search results associated with a set of low FOO historical queries previously submitted to the search engine; and

evaluating means to evaluate the term/phrase in view of terms/phrases of the second set of term clusters to identify one or more related term suggestions.

50. A computing device as recited in claim 49, and further comprising:

calculating means to calculate that there is no match between the term/phrase and term(s)/phrase(s) from the first set of term clusters, the first set being based on high FOO historical queries; and

responsive to the calculating, identifying means to identify a match between the term/phrase and term(s)/phrase(s) from one or more of the second set of term clusters, the second set being based on low FOO historical queries; and

responsive to identifying, generating means to generate related term suggestion(s) comprising the term(s)/phrase(s).